X-520-65-274

NASA TM X-55313

OPTIMUM QUANTIZATIONS

BY

J. C. MORAKIS

JULY 1965

GODDARD SPACE FLIGHT CENTER

GREENBELT, MARYLAND

X-520-65-274

OPTIMUM QUANTIZATIONS

BY
J. C. MORAKIS

July 1965

OPTIMUM QUANTIZATIONS

Abstract

11209

The transformation required by the quantization process of a continuous variable results in an error that may be considered as noise. We define Optimum Quantization as the type that minimizes the mean square (or peak) error for a constant number of quanta. This optimization is possible when the probability density function of the continuous variable is known.

Relations involving the quantization intervals as a function of the probability density function and the number of quanta are developed for three cases, i.e., when the mean of a quantum represents the values of the variable in that interval after quantization, when the midpoint assumes the above representation and when the probability density function is approximated by a piecewise uniform probability density function with discontinuities at end points of the intervals. In the last case, the mean and the midpoint for each interval coincide.

Interesting results are obtained for the special case when the continuous variable has a uniform probability density function; optimization of this case results in uniform quantization which incidentally produces the maximum mean square quantization error.

I. INTRODUCTION

The majority of telemetry systems is concerned with the measurement and transmission of information about the value of a finite range continuous variable.

A typical telemetry system such as the one shown in Figure 1 exposes some of the most important transformations that the variable is usually subjected to, from the point of measurement (or source) to the recording station (or receiver).

At the receiver, the detected value of the variable may be in error due to errors, or equivalently noise, introduced by the transformation and the medium.

There are always errors in the received value of the variable whenever this variable goes through an irreversible transformation.* For example, referring to the building blocks of Figure 1, sampling is not an irreversible transformation if the sampling rate is more than or equal to $1/2W$ where W is the highest signal frequency, or in some cases the bandwidth. However, quantization and addition of noise are irreversible transformations resulting in errors.

Although the statistics of the errors due to individual disturbances can be easily found, in case of more than one disturbance, these disturbances must be properly combined. The two most serious disturbances referred to here are the error due to quantization and the channel noise.

*An irreversible transformation is defined here as a transformation T which, if applied to x, will result in xT and whose inverse cannot restore x by the operation $(xT)T^{-1}$, or if T has no inverse.

To provide the necessary background for this discussion, the second part of this paper is devoted to the definitions and the types of quantization errors and their strong correspondence to the choice of the point that represents the infinite number of points of a quantum. This part ends with a proof that for uniform quantization the error is independent of the probability density function of x.

Part III is devoted to an introduction of information measure, and the information loss due to quantization and gaussian noise. This part is concluded with a comparison of the errors due to channel noise and those due to quantization with a brief description of their dependence on n, the number of levels. In addition, the two disturbances are combined for an optimum choice of n, the number of quantization levels.

Part IV utilizes the probability density function of a random variable x to optimize the quantization process by minimizing the quantization error, e_q, under three different conditions.

Finally, Part V is devoted to discussion and some results obtained by using typical probability density functions as examples.

II. DEFINITIONS OF QUANTIZATION ERRORS

Quantization is a many-to-one transformation of a set of values of a continuous variable into one value v_i. This transformation maps an infinite number of elements, representing the values of a continuous variable in a certain range r_i, to one element that corresponds to the range r_i.

Before proceeding any further, we must identify this element that corresponds to r_i in such a manner so as to enable ourselves to define a measure on the quantization error.

Since r_i, being an abstract set of points, cannot be represented by one value, two alternate values that could be used in lieu of r_i are the mean of x in the interval r_i and the midpoint of this interval.

The mean value of x in the interval (x_{i-1}, x_i) is expressed by:

$$m_i = \frac{\int_{x_{i-1}}^{x_i} xp(x)dx}{\int_{x_{i-1}}^{x_i} p(x)dx} \tag{1}$$

In this case, the average error is:

$$\langle e \rangle = \int_{r_i} (x-m_i)\, p(x)dx = 0 \tag{2}$$

The midpoint of the interval (x_{i-1}, x_i) is defined as:

$$y_i = \frac{x_{i-1} + x_i}{2} \tag{3}$$

In this case, the mean error is not generally equal to zero. However, the peak error becomes $\pm q_i/2$ where q_i is the ith quantization interval;

$$q_i = x_i - x_{i-1}$$

The significance of these two types of characterizations of r_i will become evident by referring to Figure 2a where the random variable, x, has been uniformly quantized into six intervals x_o, x_1, -----, x_6. If the range

of x, r_i, is identified with the bottom of each interval x_{i-1}, the error is plotted in Figure 2b and varies from zero to q_i.

Let us now draw a set of lines of value v_i on the error curve (Figure 2b) and call them the zero error lines for each interval. This procedure is equivalent to identifying the ranges r_i by $x_{i-1} + v_i$; the error in this case will vary from $-v_i$ to $q_i - v_i$.

The peak error is either $|-v_i|$ or $q_i - v_i$ whichever is larger. By letting

$$|-v_i| = q_i - v_i \quad \text{or} \quad v_i = q_i/2$$

we minimize the peak error. Thus by setting v_i equal to the midpoint we minimize the peak error.

Similarly, the mean square error is

$$\langle e^2 \rangle = \overline{(x - v_i)^2} = \overline{x^2} - 2\bar{x} v_i + v_i^2$$

To find the value of v_i for minimum square error, we differentiate the above expression with respect to v_i and set it equal to zero.

$$\frac{d}{dv_i} \langle e^2 \rangle = -2\bar{x} + 2 v_i = 0 \quad \text{or} \quad v_i = \bar{x} = m_i$$

So the mean square error is minimum when the r.v.x is identified with the mean of the i^{th} quantum.

In summarizing, we have described here two methods of identifying the quantized variable x in the i^{th} quantum.

(a) Letting $x = m_i$; $x_{i-1} < x < x_i$, we minimize the mean square error in that interval.

(b) Letting $x = y_i$; $x_{i-1} < x < x_i$, we minimize the peak error in the i^{th} interval.

When the probability density function within an interval, $p_i(x)$ is uniform as shown in Figure 3, then $p_i(x)$ is constant during that interval and the mean as given by (1) is equal to the midpoint as given by (3). This implies that either type of identification of the interval r_i will result in identical optimization.

The special case of uniform quantization has been carried out in Appendix A with the following results:

The probability density function of the error generated by replacing the variable in the interval (x_i, x_{i+1}) by its mean or midpoint is uniform with

$$p(e) = \frac{1}{q} \quad \text{for} \quad -q/2 < x < q/2$$

$$= 0 \quad \text{otherwise}$$

resulting in zero mean error, q/2 peak error and $\overline{e^2} = \overline{e_i^2} = q^2/12$ independently of the overall probability density function of the random variable x.

III. INFORMATION LOSS DUE TO QUANTIZATION

The information conveyed by a message is the amount of new knowledge acquired by receiving the message. For example, if we knew the contents of the message before reception, its reception would not convey any information.

The measure of information in the binary system could be defined as the minimum number of yes and no answers necessary to identify one out of n source symbols*, and the units are bits (for binary digits). Thus, the self-information of a discrete signal is

$$H(x) = -\sum_{n} p(x_j) \log p(x_j) \quad (4)$$

**

and it becomes maximum when all $p(x_j)$ are equal***, in this case:

$$H(x) = -\log p(x_j) \sum_{n} p(x_j) = -\log p(x_j) \quad (5)$$

For the continuous case, the situation is somewhat similar with p(x) being a uniform distribution and $p(x_j)$ substituted by P(x = X). Obviously, the latter expression is a constant for all x and approaches zero. Application of (5) for this case yields:

$$H(x) = \log \left(\frac{1}{P(x = X)} \right) \quad (6)$$

*An equivalent definition is the number of bits necessary to specify any number less than or equal to n in binary form.
**Log x = $\log_2$ x unless otherwise specified.
***Fano. (1)

This expression goes to infinity as $P(x = X)$ goes to zero. The result is very logical and it means that we need an infinite number of digits to represent the value of an analog variable. The usual practice is to give the value of the variable to a certain decimal point, a process equivalent to assigning the values of the variable in n ranges and specifying the range in which the value lies. This process is called quantization.

The self information of a quantized signal is:

$$H(x) = \sum_{j=1}^{n} P(x_{j-1} < x < x_j) \log \frac{1}{P(x_{j-1} < x < x_j)} \tag{7}$$

The minimum self-information occurs when $P(x_{j-1} < x < x_j)$ is constant for all j, and in this case:

$$P(x_i < x < x_{i+1}) = \int_{x_i}^{x_{i+1}} p(x)dx = \frac{1}{n}$$

and $$H(x) = \log n$$

n is the number of quanta and is finite, resulting in a finite H(x).

We have seen in the last paragraph how the information of a continuous variable is reduced by the quantization process from an infinite to a finite value. The information loss seems to be infinite irrespective of the value of n, a finite number; however, if we re-examine our discussion, we shall discover that in real life the measurement of a variable cannot possibly contain infinite information due to many limitations, two of the most

important being the accuracy of the measuring apparatus and the system* noise. In general, the system noise is the main factor in determining the number of distinguishable levels of an analog value. For example, if the average signal power is S and the average noise (gaussian) power is N, the number of distinguishable signal levels is $\sqrt{\frac{S}{N}}$ and if the zero value is added the number of distinguishable voltage levels becomes $\sqrt{\frac{S}{N} + 1}$. Applying equation (6), the information becomes

$$H(x) = \frac{1}{2} \log \left(\frac{S+1}{N} \right) \text{ bits}$$

Although this is an unconventional method of deriving the information for a given S/N the answer is correct.

If the signal has a bandwidth W and a duration T, invoking the sampling theorem, the signal may be (theoretically) described by samples spaced 1/2W seconds apart; thus, the number of samples in an interval T will be equal to 2WT. The resulting maximum information per interval T is:

$$2WT \; 1/2 \log (S/N + 1)^{**}$$

and the maximum information rate (or channel capacity) is:

$$W \log (S/N + 1)$$

The obvious deduction from the above formula is that there is no signal that contains infinite information unless the noise is zero or S/N

*The word system may refer to anything from an amplifier to a transmitter-channel-receiver configuration.

**When S/N = 1 H_{max} = WT the dimension of the signal space.

is infinite, an impossible event in physical systems; consequently, the loss of information due to quantization is finite and may be reduced by increasing the number of quanta (or levels). An increase in the number of levels will decrease the quantization error at the transmitter, but will increase the probability of error at the receiver, because the signal corresponding to one level $(\text{range}/n)^2$ decreases as n increases, thus decreasing the received S/N.

Referring to Figure 1, the quantization mean square error is an error that occurs in the transition* x to x_i and is given by:

$$\overline{e_q^2} = \frac{q^2}{12} = \frac{R^2}{12n^2}$$

The error at the receiver due to channel noise is an error due to the transition from x_i to y_j $(i \neq j)$, and it is expressed in a statistical manner, because in quantized signal transmission an error will be effective only if it causes the signal to change level. Consequently, for N gaussian, no matter how small N is in comparison to R/n, the fact that N can assume values of $-\infty$ to ∞ imposes a probability that the signal will change level. For quantized signals with n levels of unequal length (nonuniform) the error produced by going from level i to level j is e_{ij} associated with a

*Uniform quantization is assumed.

probability P_{ij} that depends on the distance between levels i and j and the type of modulation. The mean square error is then:

$$\overline{e_N^2} = \sum_i \sum_j e_{ij}^2 P_{ij} \tag{11}$$

Solving for $\overline{e_N^2}$ for the typical case of PCM (uniform quantization into 2^k levels and binary signals), P_{ij} becomes a constant (P) and e_{ij} varies uniformly over all discrete distances which are q multiplied by powers of two; these distances are:

$$q,\ 2q,\ 2^2q,\ ---,\ 2^{k-1}q;\quad (q = R/n)$$

and (11) becomes

$$\overline{e_N^2} = \sum_{i=0}^{k-1} (q\ 2^i)^2\ P = \left(\frac{R}{N}\right)^2 \frac{2^{2k}-1}{3}\ P$$

In general, the two most significant errors, $\overline{e^2}_q$ and $\overline{e^2}_N$ are independent; thus the total error is given by (12) as

$$\overline{e^2} = \overline{e^2}_q + \overline{e^2}_N \tag{12}$$

where

$$\overline{e_q^2} = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} (x - m_i)^2\ p_i(x)dx \tag{13}$$

and $\overline{e^2}_N$ is given by (11).

Although $\overline{e^2}_q$ is independent of p(x) for the case of uniform quantization, generally, for optimum quantization*, $\overline{e^2}_q$ is a function of p(x); from (11) it is obvious that $\overline{e^2}_N$ is a function of e_{ij} and P_{ij} which are in turn functions of R, n, S/N, the type of noise, the type of modulation, and the type of detection.** Optimum quantization will involve optimization of n and q_i under the criterion of minimum mean square error as given by (12). Minimization of (12) will invariably result in an equation that could only be solved by numerical methods. To simplify the problem, let us examine e_N and e_q separately.

First it should be observed that e_N and e_q have no minimum as a function of n all other conditions being constant. However, they are both monotonic functions of n, one increasing and the other decreasing. The opposing effects of n on e_N and e_q suggest optimum operation when the two errors become equal as expressed by equation 14.

$$\overline{e_q^2} = \overline{e_N^2} \tag{14}$$

Once n is thus chosen, $\overline{e^2}_q$ is minimized by using optimum quantization with respect to the intervals q_i as shown in section IV; with the new value of $\overline{e^2}_q$ a new n is chosen and the process is repeated. With a small amount of experience one might obtain a good estimate of $\overline{e^2}_q$, which may be 20-30% less than that due to uniform quantization, calculate n and then optimize for q_i in only one trial.

*This will be proven in IV.

**For example:

P_{ij} for FSK = $P = \frac{1}{2}e^{-\frac{E}{2N_0}}$ from Lawton (2).

P_{ij} for PCM antipodal signals is an error function.

IV. OPTIMUM QUANTIZATION

It has been shown in part II that under uniform quantization the results are independent of the overall probability density function of the random variable x. It will be shown here that the quantization error may be minimized if p(x) is utilized in determining the values of the quanta q_i.

The first type of optimization will be carried out under the condition that the quantization process maps all x in the interval (x_{i-1}, x_i) onto the mean m_i or $x = m_i$ for $x_{i-1} < x < x_i$. The second type of optimization will be carried out under the condition that the quantization process maps all x in the interval (x_{i-1}, x_i) onto the midpoint y_i, and in the third type of optimization the approximation $p_i(x) = \overline{p_i(x)}$ will make $m_i = y_i$ thus forcing the above two conditions to coincide.

1. Optimum Quantization with $(x_{i-1}, x_i) \rightarrow m_i$

To obtain the optimum quantization for a given number of quanta, n, and under the condition of mapping the i^{th} interval (x_{i-1}, x_i) onto m_i, we must minimize the average mean square error as given by (B-1). This is done in Appendix B and the results are given by (B-2).

$$x_i = \frac{m_i + m_{i+1}}{2} \tag{B-2}$$

if the range of x extends from a to b (B-2) represents n-1 equations with $x_o = a$ and $x_n = b$.

2. Optimum Quantization with $(x_{i-1}, x_i) \rightarrow y_i$

In this case, we are actually defining optimum quantization as the one that minimizes the average peak error. This is carried out in Appendix C with the resultant n-1 equations for x_i given by (C-3)

$$x_{i+1} - 2x_i + x_{i-1} = \frac{A_i - A_{i+1}}{p(x_i)} \qquad \text{(C-3)}$$

or

$$q_i - q_{i+1} = \frac{A_{i+1} - A_i}{p(x_i)}$$

3. Optimum Quantization with $p_i(x) \cong \overline{p_i(x)}$

If $p_i(x)$ is approximated by $\overline{p_i(x)}$, a uniform distribution for each interval q_i, both $p_i(x)$ and m_i are functions of x_i and vary in such a manner so that the area in a quantum is preserved. Obviously this type of approximation will make $m_i = y_i = \frac{x_i + x_{i-1}}{2}$ and we do not know the points of discontinuities until the error is minimized.

From Appendix D the equations relating the points x_1, x_2 ------ x_{n-1} are from (D-6)

$$(x_i - x_{i-1})^2 [2p(y_i) + p(x_i)] = (x_{i+1} - x_i)^2 [2p(y_{i+1}) + p(x_i)] \qquad \text{(D-6)}$$

where

$$p(x_i) = p(x = x_i)$$

Another way to express this equation is:

$$2q_i A_i = 2q_{i+1} A_{i+1} + p(x_i) (q_i + q_{i+1}) (q_{i+1} - q_i) \qquad \text{(D-6)}$$

Since $q_{i+1} - q_i$ may be small, the effect of the $p(x_i)$ term on the equation is very small. So as a preliminary estimate on x_i for the solution of (D-6) one may solve

$$q_i A_i = q_{i+1} A_{i+1} \qquad (15)$$

DISCUSSION:

It has been shown here that the type of quantization resulting in the highest error is the most common one; i.e., the uniform type. The use of the probability density function of the variable in choosing the quantizing levels reduces the quantization error. This error reduction increases as the "non-uniformity" of the probability density function of x increases.

Extending the above statement the error reduction becomes zero at one extreme when the p.d.f. of x is uniform.

The term "non-uniformity" implies the difference in p.d.f. from one level to the next. Thus, an extremely non-uniform p.d.f. for a given number of levels may become more uniform as the number of levels is increased, thus decreasing the effectiveness of the methods discussed in this paper. On the other hand, the self information becomes maximum when each symbol or quantization interval is equiprobable which implies uniform distribution. Intuitively speaking, it seems that the loss of self information due to non-uniformity, is compensated by the error reduction due to advantageous optimum quantization.

Table I presents some results of improvement in quantization error by the use of optimum quantization.

Inspection of the exact and approximate distribution method results will show that for all practical purposes the approximation is reasonably valid.

APPENDIX A

UNIFORM QUANTIZATION

This type of quantization is common because of its relative simplicity of implementation. If the interval is represented by the mean, the quantization process will result in an average mean square error as given by (A-1)

$$\overline{e^2} = \sum_i \int_{x_{i-1}}^{x_i} (x-m_i)^2 \, p(x)dx \tag{A-1}$$

where m_i is given by (1)

To prove this, we must first find the mean square error in each interval.

$$\overline{e_i^2} = \int_{x_{i-1}}^{x_i} (x-m_i)^2 p_i(x)dx = \frac{1}{A_i} \int_{x_{i-1}}^{x_i} (x-m_i)^2 \, p(x)dx \tag{A-2}$$

where

$$p_i(x) = \frac{p(x)}{A_i} \; [u(x-x_{i-1}) - u(x-x_i)] \tag{A-3}$$

and u is the unit step function; A_i, the area under p(x) in the ith interval, is the normalization factor for $p_i(x)$. Averaging over all intervals, one obtains the average mean square error

$$\overline{e^2} = \sum_i \overline{e_i^2} \; P \; (x \in r_i) = \sum_i \overline{e_i^2} \; A_i \; * \tag{A-4}$$

Next, substituting (A-2) into (A-4) we obtain (A-1) QED.

A much simpler equation for the mean square error may be obtained if p(x) is approximated by a constant density function in the ith interval given by:

$$_i(x) = \frac{\int_{x_{i-1}}^{x_i} p(x)dx}{x_i - x_{i-1}} \tag{A-5}$$

---

* $P \; (x \in r_i)$ reads: the probability that x lies in the ith range;
$P \; (x \in r_i) = P \; (x_{i-1} < x < x_i)$

This approximation is illustrated in Figure A-1.

The error probability density function turns out to be uniform as shown in Figure A-2 and the average mean square error is $q^2/12$ independently of the probability density function of x. Referring to Figure A-1 the smooth line of p(x). The p.d.f. in each interval $p_i(x)$ is approximated by its average in that interval,

$$\overline{p_i(x)} = \frac{A_i}{x_i - x_{i-1}} [u(x-x_{i-1}) - u(x - x_i)] \tag{A-6}$$

and the new piecewise uniform function is the new approximate p.d.f., w(x).

$$w(x) = \sum_i^n \overline{p_i(x)} = \sum_i^n \frac{A_i}{q} [u(x - x_{i-1}) - u(x - x_i)] \tag{A-7}$$

This approximation brings the mean and the midpoint of each interval to the same point

$$m_i = y_i = \frac{x_i + x_{i-1}}{2}$$

So far we have taken a range of values of x (the infinite number of points between x_i and x_{i-1}) and have identified them by their mean or midpoint m_i or y_i respectively. The error due to this quantization is:

$$e = x - y_i \quad \text{or} \quad x = e + y_i \tag{A-8}$$

For y_i and e independent:

$$M_x = M_{y_i} M_e \ * \tag{A-9}$$

---

* M is the moment generating function of the random variable.

Since the moment generating function is the Fourier transform of the p.d.f. (A-9) may be rewritten as:

$$\mathcal{F}\{p(x)\} = \mathcal{F}\{p(y_i)\}\ \mathcal{F}\{p(e)\} \tag{A-10}$$

Equation (A-9) implies that

$$p(x) = p(y_i) \otimes p(e) \ ^{(1)} \tag{A-11}$$

If p(x) is approximated by w(x), the piecewise uniform curve, (A-11) may be rewritten as:

$$w(x) = p(y_i) \otimes p(e) \tag{A-12}$$

At this point, it should be noticed that $p(y_i)$ is a train of delta functions centered at $x = y_i$ and of area A_i. If the convolution of p(e) and a train of delta functions $p(y_i)$ is to give the function w(x) of Figure A-1, it is clear that p(e) must be a rectangular pulse function as shown in Figure A-2.

Normalization of p(e) results in $K = 1/q$. The mean is equal to the midpoint (uniform distribution) which is obviously zero. So the mean error is zero from

$$\overline{e_i} = \int_{q/2}^{q/2} \frac{1}{q}\ ede = 0$$

and the mean square error per interval is:

$$\overline{e_i^2} = \sum_{i}^{n}\ e^2\ \frac{1}{q}\ de = q^2/12 \tag{A-13}$$

Averaging this error over all the intervals

$$\overline{e^2} = \sum_{i}^{n}\ \overline{e_i^2}\ P(y_i) = \overline{e_i^2} \sum_{i}\ P(y_i) = \overline{e_i^2} = q^2/12 \tag{A-14}$$

---

(1) $\otimes$ is the symbol indicating convolution.

APPENDIX B

OPTIMUM QUANTIZATION WITH $(x_{i-1}, x_i) \rightarrow m_i$

Although calculus of variations could be used to minimize the error, the procedure to be followed here will be a simple one of taking the partial derivative of the error with respect to each variable x_i at a time, holding the remaining variables constant. For n quanta there are n-1 variables and the above process will yield n-1 equations with n-1 unknowns, when the partial derivatives are set equal to zero.

If the range of x extends from a to b and the number of quanta is n, the expression for the average mean square error is:

$$\overline{e^2} = \int_a^{x_1} (x-m_1)^2\, p(x)dx + \int_{x_1}^{x_2} (x-m_2)^2\, p(x)dx + - - - - - -$$

$$+ \int_{x_{i-1}}^{x_i} (x-m_i)^2\, p(x)dx + - - - - - - - + \int_{x_{n-1}}^{b} (x-m_n)^2\, p(x)dx \qquad \text{(B-1)}$$

Differentiating with respect to x_i

$$\frac{\partial \langle e^2 \rangle}{\partial x_i} = 0 + 0 + \cdots\cdots + \frac{\partial}{\partial x_i} \int_{x_{i-1}}^{x_i} (x-m_i)^2\, p(x)dx + \int_{x_i}^{x_{i+1}} (x-m_{i+1})^2 p(x)dx + 0 + 0 - - + 0$$

$$= (x_i - m_i)^2\, p(x_i) - (x_i - m_{i+1})^2\, p(x_i)$$

Setting the above equation equal to zero for all i

$$(x_i - m_i)^2 = (x_i - m_{i+1})^2 \quad \text{if} \quad p(x_i) \neq 0$$

The above equation yields

$$x_i = \frac{m_i + m_{i+1}}{2} \qquad \text{(B-2)}$$

APPENDIX C

OPTIMUM QUANTIZATION WITH $(x_{i-1}, x_i) \rightarrow y_i$

For the conditions imposed here we must minimize the average peak error. The peak error e_p is

$$e_p = \frac{x_i - x_{i-1}}{2} = q_i/2 \tag{C-1}$$

and the average peak error is

$$\overline{e_p} = \frac{1}{n} \sum_i \frac{q_i}{2} P(x \in q_i) = \frac{1}{n} \sum_i q_i \int_{x_i}^{x_i} p(x)dx \tag{C-2}$$

By differentiating (C-2) with respect to x_i and by setting equal to zero we obtain the condition of optimum quantization for peak error which is given by C-3.

$$2 x_i - x_{i+1} - x_{i-1} = \frac{A_{i+1} - A_i}{p(x_i)} \tag{C-3}$$

Equation (C-3) represents n-1 equations.

APPENDIX D

OPTIMUM QUANTIZATION USING THE APPROXIMATE pdf

The approximation given by $p_i(x) = \overline{p_i(x)}$ * will change the dotted line pdf of Figure B-1 to the solid line pdf. In addition, the mean m_i equals the midpoint y_i which is equal to one-half of the sum of the limit points of the interval.

The method of optimization will be similar to the one employed in Appendix B with differentials replaced by differences because of the nature of $\overline{p_i(x)}$. The error is given by:

$$e^2 = \sum_j \langle e_j^2 \rangle \, P(x \epsilon r_j) = \sum_j \int_{j \text{ interval}} (x-mj)^2 \, \overline{p_j(x)} \, dx$$

$$= \left. \begin{aligned} &\int_{x_o}^{x_1} (x-m_1)^2 \, \overline{p_1(x)} dx + \int_{x_1}^{x_2} (x-m_2)^2 \, \overline{p_2(x)} \, dx + \text{----} \int_{x_{i-1}}^{x_i} (x-m_i)^2 \, \overline{p_i(x)} \, dx \\ &+ \int_{x_i}^{x_{i+1}} (x-m_{i+1})^2 \, \overline{p_{i+1}(x)} \, dx + \text{-----} + \int_{x_{n-1}}^{x_n} (x-m_n)^2 \, \overline{p_n(x)} \, dx \end{aligned} \right\} \text{(D-1)}$$

If x_i is allowed to vary by a small amount Δx while the rest of the limits are kept constant, the new ms error $\overline{e^2}^1$ is given by (D-1) with the following changes in the ith and i + 1st integrals. The limit x_i is replaced by $x_i + \Delta x$,

$$\overline{p_i(x)} \rightarrow \overline{p_i(x)}^1, \quad \overline{p_{i+1}(x)} \rightarrow \overline{p_{i+1}(x)}^1, \quad m_i \rightarrow m_i^1, \text{ and } m_{i+1} \rightarrow m_{i+1}^1$$

The variation of the ms error is given by the difference of $\overline{e^2}^1$ and $\overline{e^2}$

---

* See (A-5) and (A-6) for definition.

$$\Delta(\overline{e^2}) = \overline{e^2}^1 - \overline{e^2} = \int_{x_{i-1}}^{x_i+\Delta x} (x-m_i^1)^2 \overline{p_i(x)}^1 dx + \int_{x_i+\Delta x}^{x_{i+1}} (x-m_{i+1}^1)^2 \overline{p_{i+1}(x)}^1 dx$$

$$- \int_{x_{i-1}}^{x_i} (x-m_i)^2 \overline{p_i(x)}\, dx - \int_{x_i}^{x_{i+1}} (x-m_{i+1})^2 \overline{p_{i+1}(x)}\, dx \qquad \text{(D-2)}$$

Using the fact that all pdf's in (D-2) are uniform and m is half of the sum of the limits, i.e., $m_i = \frac{x_{i-1} + x_i + \Delta x}{2}$, ----- , $m_{i+1} = \frac{x_i + x_{i+1}}{2}$

(D-2) may be rewritten as*:

$$\frac{1}{12}\left[\overline{p_i(x)}^1 (x_i+\Delta x - x_{i-1})^3 - \overline{p_i(x)}\ (x_i - x_{i-1})^3 + \overline{p_{i+1}(x)}^1 (x_{i+1}-x_i-\Delta x)^3 - \overline{p_{i+1}(x)}\ (x_{i+1} - x_i)^3\right] \qquad \text{(D-3)}$$

To simplify (D-3) let us refer to Figure D-2 and consider the integral that represents the area under p(x) in the intervals adjoining $x_i + \Delta x$

$$\int_{x_{i-1}}^{x_i+\Delta x} (p(x)dx = \int_{x_{i-1}}^{x_i} p(x)dx + \int_{x_i}^{x_i+\Delta x} p(x)\, dx \qquad \text{(D-4)}$$

similarly

$$\int_{x_i+\Delta x}^{x_{i+1}} p(x)dx = \int_{x_i}^{x_{i+1}} p(x)dx - \int_{x_i}^{x_i+\Delta x} p(x)dx \qquad \text{(D-5)}$$

Substituting (A-5) into (D-4) and (D-5) and letting $\int_{x_i}^{x_i+\Delta x} p(x)dx = \Delta A$

we get:

$$(x_i+\Delta x - x_{i-1})\ \overline{p_i(x)}^1 = (x_i - x_{i-1})\ \overline{p_i(x)} + \Delta A \qquad \text{(D-4)}$$

$$(x_{i+1} - x_i - \Delta x)\ \overline{p_{i+1}(x)}^1 = (x_{i+1} - x_i)\ \overline{p_{i+1}(x)} - \Delta A \qquad \text{(D-5)}$$

*if $m = \frac{a+b}{2}$, $\int_a^b (x-m)^2 dx = \frac{(b-a)^3}{12}$

Substituting (D-4) and (D-5) into (D-3) and expanding

$$\Delta \overline{e^2} = \frac{1}{12} \left\{ (x_i - x_{i-1}) \overline{p_i(x)} \left[2 (x_i - x_{i-1}) \Delta x + (\Delta x)^2 \right] \right.$$

$$+ (x_{i+1} - x_i) \overline{p_{i+1}(x)} \left[-2 (x_{i+1} - x_i) \Delta x + (\Delta x)^2 \right]$$

$$\left. + \Delta A \left[(x_i - x_{i-1})^2 - (x_{i+1} - x_i)^2 + 2 (x_{i+1} - x_{i-1}) \Delta x \right] \right\}$$

when $\Delta x \rightarrow 0 \qquad \Delta A \rightarrow p(x_i) \Delta x$

and:

$$\lim_{\Delta x \to 0} \frac{\Delta \overline{e^2}}{\Delta x} = \frac{1}{12} \left\{ 2 (x_i - x_{i-1})^2 \overline{p_i(x)} - 2 (x_{i+1} - x_i)^2 \overline{p_{i+1}(x)} + \right.$$

$$\left. p(x_i) \left[(x_i - x_{i-1})^2 - (x_{i+1} - x_i)^2 \right] \right\}$$

setting the derivative equal to zero and replacing $x_j - x_{j-1}$ by q_j

we obtain the relation for optimum quantization

$$2 q_i^2 \overline{p_i(x)} = 2 q_{i+1}^2 \overline{p_{i+1}(x)} + p(x_i) (q_{i+1}^2 - q_i^2) \qquad \text{(D-6)}$$

Repeating this for all i is equivalent to considering (D-6) as n-1 equations.

These n-1 equations together with

$$\sum_{i=1}^{n} q_i = x_n - x_o$$

form the n equations with n unknowns.

GLOSSARY

x A continuous random variable (r.v.)

x_i The limits of the quantizing intervals.

y_i The midpoint of the ith interval $\left(y_i = \frac{x_i + x_{i-1}}{2}\right)$

m_i The mean of the ith interval $\left(m_i = \int_{x_{i-1}}^{x_i} x\, p(x)dx \Big/ \int_{x_{i-1}}^{x_i} p(x)dx\right)$

r_i The ith range of x $(x_{i-1} < x < x_i)$

R The range of x (also the value of the range = $x_n - x_o$)

n The number of quantization levels.

q_i The value of the ith quantization interval $(q_i = x_i - x_{i-1})$

$p(x)$ The probability density function (pdf) of x

$p_i(x)$ The normalized pdf of x in the ith interval $\left(p_i(x) = \frac{p(x)}{A_i}\ ;\ x_{i-1} < x < x_i\right)$

$p(x_i) = p(x=x_i)$

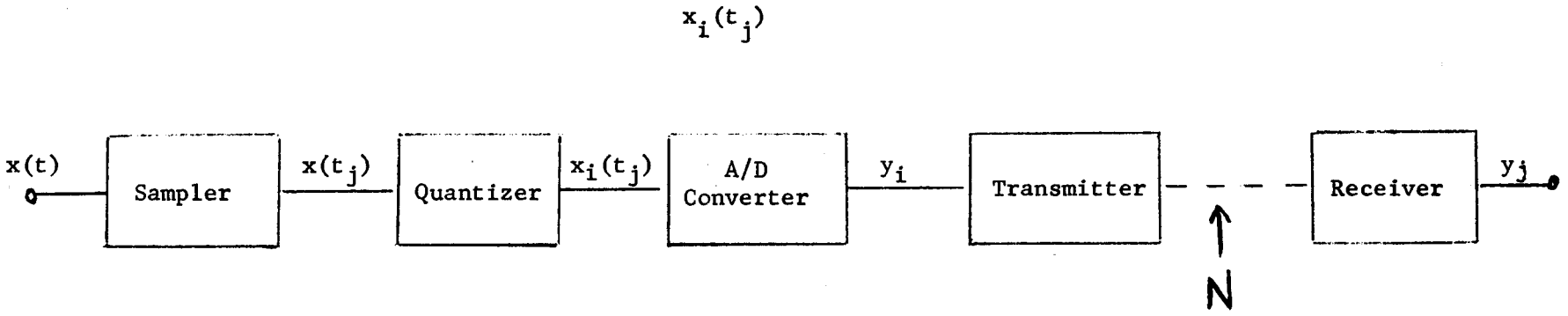
$p(y_i)$ The probability that x is represented by y_i an impulse function located at $x = y_i$ with weight $A_i = P(x_{i-1} < x < x_i) = P(x \in r_i)$

$\overline{p_i(x)}$ = the mean pdf in the ith interval $\left(p_i(x) = \left(\int_{x_{i-1}}^{x_i} p(x)dx\right) \Big/ (x_i - x_{i-1})\right)$

S The average signal power

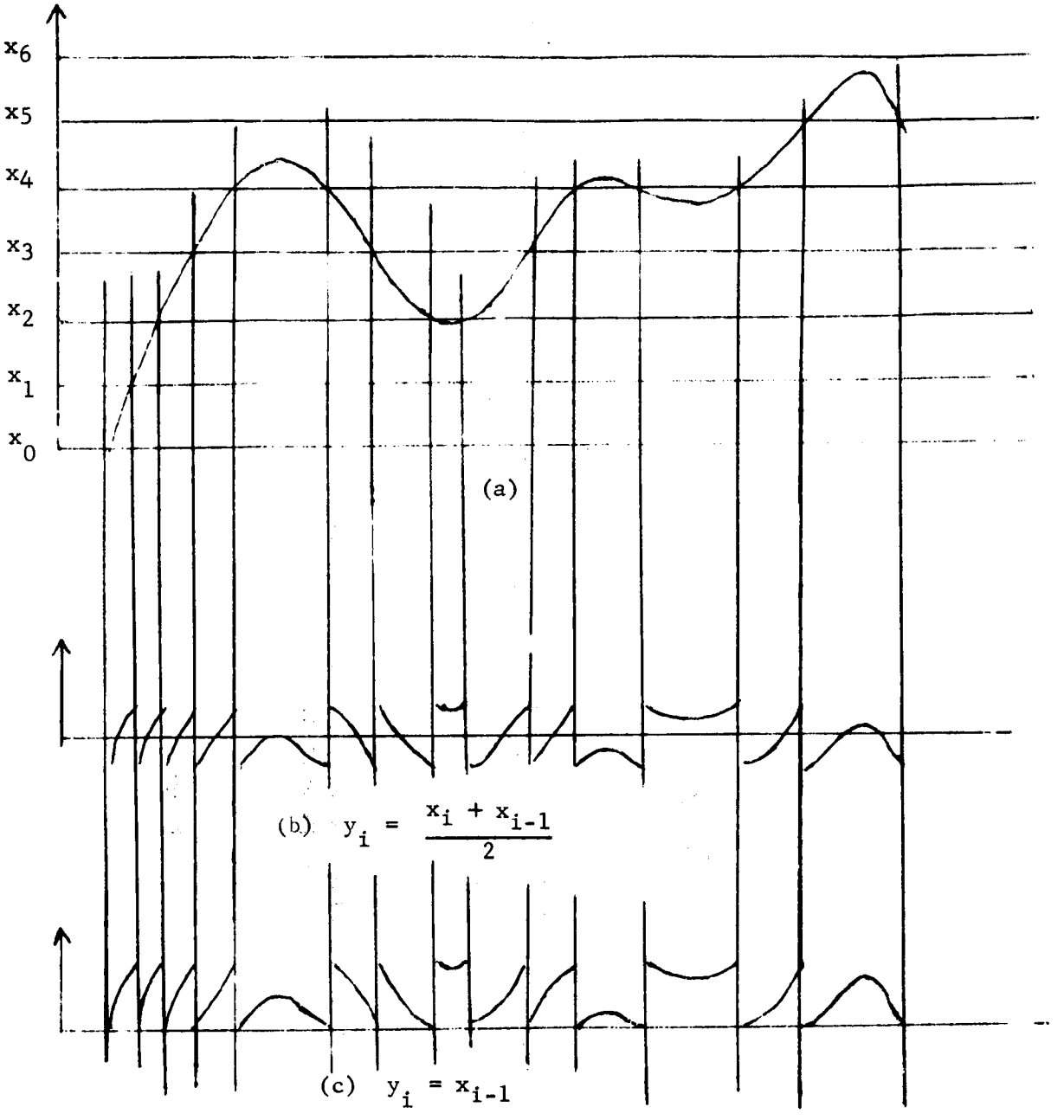
N The average thermal noise power

W The signal bandwidth



A Transmission System employing sampling
and quantization

Figure 1



Quantization of a continuous function and the error with (b) midpoint representation and (c) unbiased representation.

Figure 2

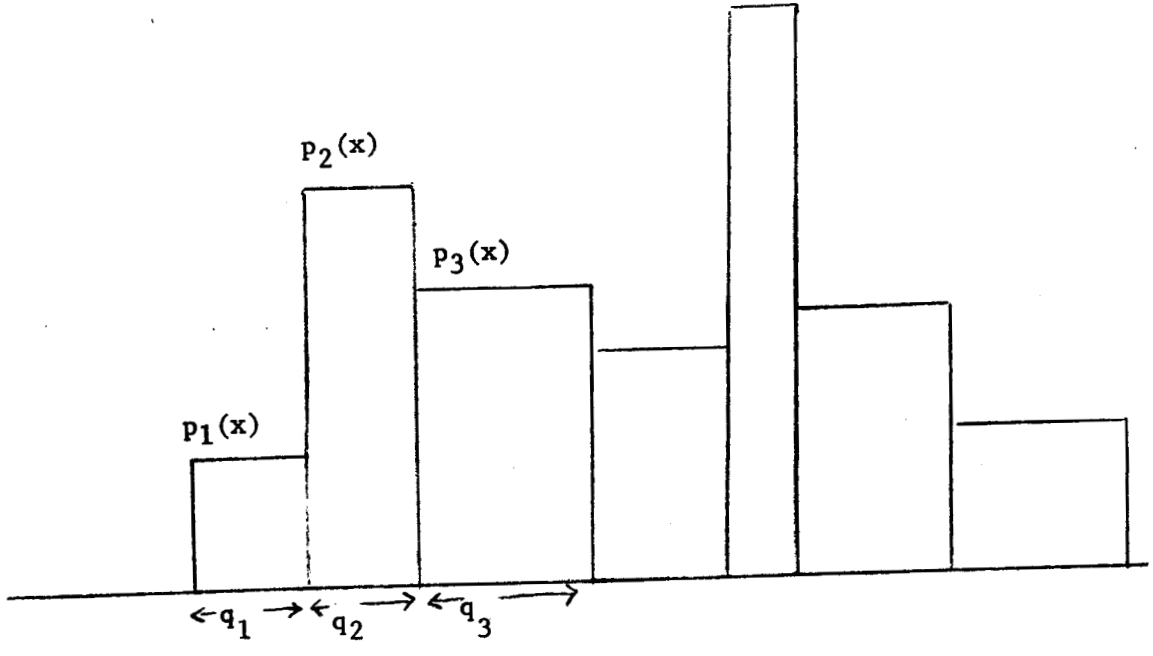
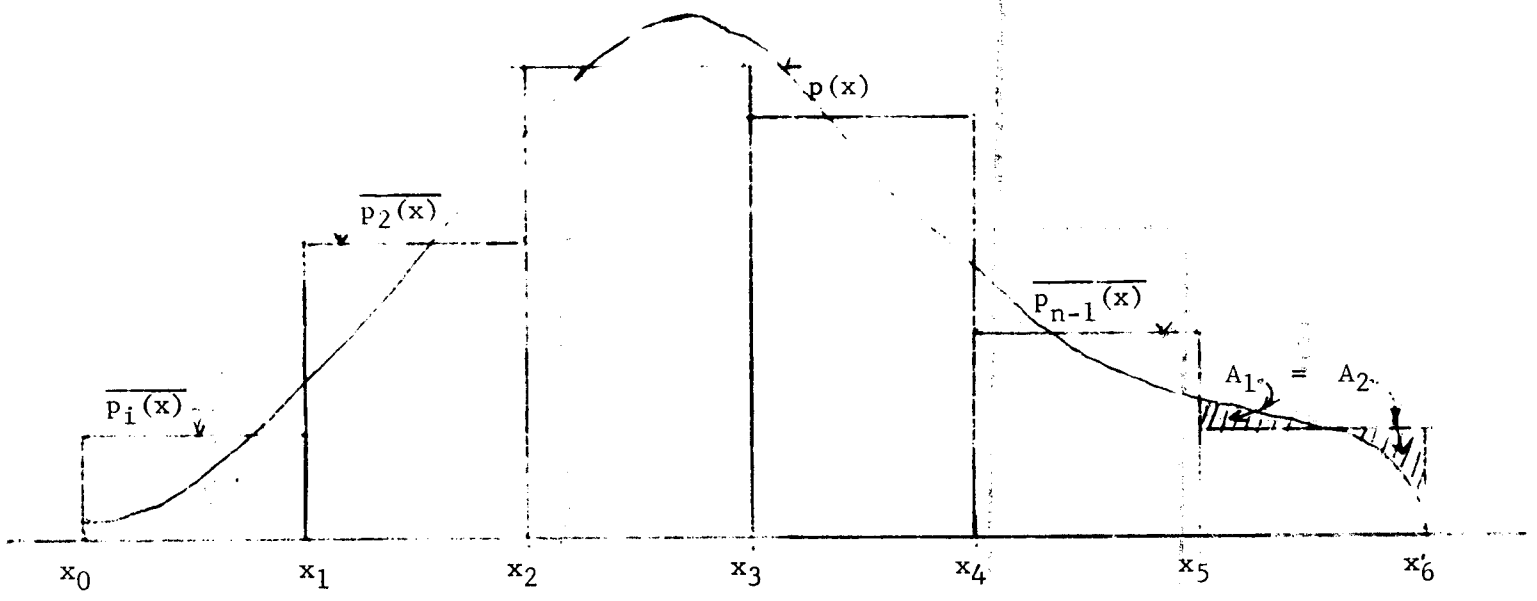
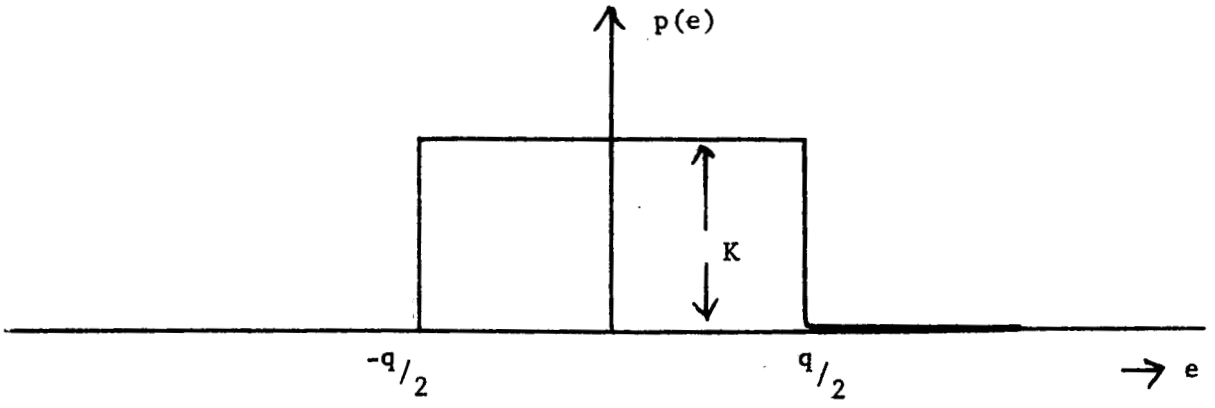


Figure 3



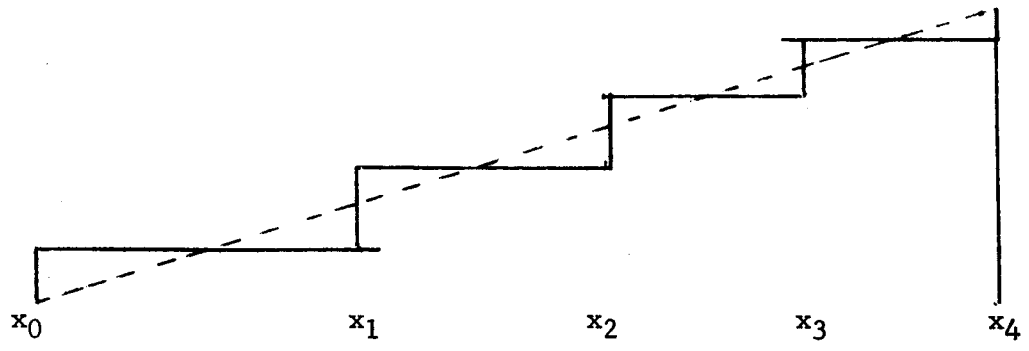
The exact and approximate distribution functions

Figure A-1



The pdf of the uniform quantization error.

Figure A-2



Exact (dotted line) and approximate pdf

Figure D-1

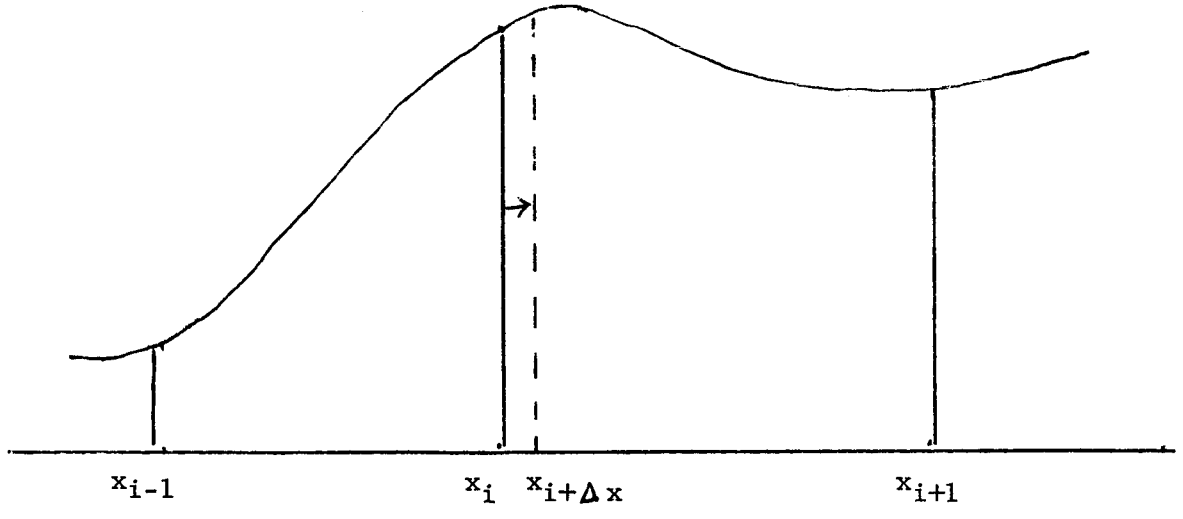


Figure D-2

p(X)	n	UNIFORM $\bar{e}^2$	OPTIMUM X_1	X_2	X_3	X_4	$\bar{e}^2$	%Improvment	OPTIMUM (APPROXIMATE) X_1	X_2	X_3	X_4	$\bar{e}^2$	%Improvment
aX (linear)	2	$.021b^2$	.618b	b			$.015b^2$	34.6	.577b	b			$.016b^2$	31.3
	4	$.005b^2$	.376b	.608b	.812b	b	$.004b^2$	29.6	.338b	.585b	.802b	b	$.004b^2$	21.7
(normal)	4	.188*	.99	∞			.118	59.6	1.157	∞			.153	22.4
	8	.047*							.523	1.1	1.81	∞	.035	34.7
aX^2 (square)	2	$.021b^2$	.691b	b			$.011b^2$	89.0	.621b	b			$.012b^2$	72.7
	4	$.005b^2$	.469b	.678b	.85b	b	$.003b^2$	76.2	.397b	.639	.833b	b	$.003b^2$	60.5
c (uniform)	n	$\frac{1}{12n^2}b^2$	b/n	2b/n	ib/n	b	$b^2/12n^2$	0	Same.					0

NOTE: The Range is the interval (0,b)

* Truncated at ±3

TABLE I